
저자 (Authors)	이영남, 김상욱 Youngnam Lee, Sang-Wook Kim
출처 (Source)	정보과학회논문지 44(12) , 2017.12, 1333-1337(5 pages) Journal of KIISE 44(12) , 2017.12, 1333-1337(5 pages)
발행처 (Publisher)	한국정보과학회 The Korean Institute of Information Scientists and Engineers
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07319381
APA Style	이영남, 김상욱 (2017). 추천 시스템에서의 데이터 임퓨테이션 분석. 정보과학회논문지, 44(12), 1333-1337
이용정보 (Accessed)	한양대학교 166.***.182.218 2020/09/08 10:05 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

■ (네이버 후원) 2017년도 학생논문 경진대회 수상작

추천 시스템에서의 데이터 임putation 분석 (Analysis of Data Imputation in Recommender Systems)

이 영 남 [†] 김 상 욱 ^{**}
(Youngnam Lee) (Sang-Wook Kim)

요 약 추천 시스템이란 사용자가 좋아할만한 개인화된 상품을 사용자에게 제안하는 것이다. 최근 상품 수의 증가로 추천 시스템의 중요성이 날로 커지고 있지만, 데이터 희소성 문제는 여전히 추천 시스템의 대표적인 문제로 남아있다. 데이터 희소성 문제는 사용자가 전체 상품 중 일부의 상품에만 평점을 부여하여, 사용자와 상품 관계를 정확히 이해하기 힘든 것을 말한다. 이를 해결하기 위해 가장 여러 가지 접근법이 있는 그 중 대표적인 것인 데이터 임putation이다. 데이터 임putation은 사용자가 평가하지 않은 상품의 평점을 추론해 평점 행렬에 채우는 방법이다. 하지만 기존 데이터 임putation 방법은 사용자가 평가하지 않은 상품에 대한 몇 가지 특성을 놓치고 있다. 본 논문에서는 기존 방법의 한계점을 정의하고, 이를 개선하는 방안 3가지를 제안한다.

키워드: 추천 시스템, 협업 필터링, 데이터 희소성, 데이터 임putation

Abstract Recommender systems (RS) that predict a set of items a target user is likely to prefer have been extensively studied in academia and have been aggressively implemented by many companies such as Google, Netflix, eBay, and Amazon. Data imputation alleviates the data sparsity problem occurring in recommender systems by inferring missing ratings and adding them to the original data. In this paper, we point out the drawbacks of existing approaches and make suggestions for data imputation techniques. We also justify our suggestions through extensive experiments.

Keywords: Recommender system, Collaborative filtering, Data sparsity, Data imputation

· 이 성과는 2017년도 미래창조과학부의 재원에 의한 한국연구재단의 지원 (No. NRF-2017R1A2B3004581)과 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업 (IITP-2017-2013-0-00881)의 지원을 받았습니다.

[†] 비 회 원 : 한양대학교 컴퓨터소프트웨어학부 연구원
dudska0430@hanyang.ac.kr

^{**} 종신회원 : 한양대학교 컴퓨터소프트웨어학부 교수
(Hanyang Univ.)
wook@hanyang.ac.kr
(Corresponding author)

논문접수 : 2017년 5월 31일
(Received 31 May 2017)

논문수정 : 2017년 9월 1일
(Revised 1 September 2017)

심사완료 : 2017년 10월 24일
(Accepted 24 October 2017)

Copyright©2017 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 제44권 제12호(2017. 12)

1. 서 론

추천 시스템이란 사용자가 좋아할만한 개인화된 상품을 사용자에게 제안한다. 최근, 상품 수의 급격한 증가로 사용자는 자신이 가장 선호할 상품을 찾기 힘들다. 사용자가 가장 선호할 상품을 제공하는 것은 쉽지 않은 일이지만, 그것이 가능하다면 고객의 만족도와 충성심을 높일 수 있다. 그렇기 때문에 추천 시스템의 중요성은 시간이 지날수록 더 커지고 있고, Google, Ebay, Amazon, Netflix 등과 같은 회사들이 추천 시스템을 적극적으로 활용하는 이유이다. 협업 필터링은 추천 시스템에서 가장 성공적인 방법 중 하나이며, 널리 사용되고 있다. 협업 필터링은 과거에 사용자가 남긴 평점 정보를 기반으로 사용자와 상품 간의 관계를 학습한다. 협업 필터링의 목적은 2가지이다. 사용자가 부여하지 않은 평점 값을 추정하는 (1) 평점 예측과 사용자가 상대적으로 선호할 상품을 찾아내는 (2) 선호 상품 추천. 실제 환경에서는 선호 상품 추천이 더 유용하게 활용되기 때문에 본 연

구는 이 문제에 초점을 맞추어 연구[1,2,5]를 진행했다.

사용자-상품 평점 행렬 R 의 상당 부분은 비어있다. 왜냐하면 대부분의 사용자는 전체 상품 중 극히 일부의 상품만 이용하고, 이용한 상품에 한해서 평점을 부여하기 때문이다. 협업 필터링은 사용자가 남긴 평점들의 분석을 기반으로 하는데, 이와 같이 데이터가 희소한 환경에서는 사용자와 상품의 관계를 정확히 이해하기 힘들고 결과적으로 추천시스템의 정확도가 나이지게 된다. 이 문제를 **데이터 희소성 문제**라고 부르며 협업 필터링이 안고 있는 대표적인 문제이다. 이러한 데이터 희소성 문제를 해결하기 위해 신뢰 네트워크 정보 이용[8], 크라우드소싱[11], 데이터 임putation[3,12] 등 다양한 방법이 추천 시스템 분야에서 연구 되어 왔다.

데이터 임putation은 비어있는 평점에 대해 존재하는 평점만을 이용하여 사용자가 해당 상품에 부여할 것 같은 평점을 추론한, 이를 행렬 R 에 채우는 방법이다. 평점 행렬 외 신뢰 네트워크 크라우드소싱 등의 추가적인 정보를 요구하는 다른 기법과는 달리 **평점 행렬만을 이용한다는 점에서 활용의 범위가 크다**. 데이터 임putation은 추론한 평점을 추가적으로 사용하기 때문에 기존의 평점 행렬 R 을 이용하는 것보다 더 높은 추천 정확도를 보일 것으로 기대한다. 그러나 데이터 임putation은 사용자가 직접 부여한 것이 아니라 존재하는 평점을 이용해 추론한 것이기 때문에 사용자가 직접 부여할 값과의 **오차가 존재한다**. 그러므로 이 오차를 줄이는 것이 임putation 성능의 핵심요소이다.

본 연구에서는 기존 데이터 임putation 방법이 놓치고 있던 한계점 3가지를 정의하고, 이를 개선할 수 있는 3가지 방법을 제안한다. 이 3가지 방법은 협업 필터링과는 무관하기 때문에 어떤 데이터 임putation 방법에도 적용이 가능하다. 또한 실험을 통해 3가지 방법을 적용했을 때 정확도 상승이 있음을 보인다.

2. 관련 연구

기존 데이터 임putation 연구는 크게 두 가지로 나눌 수 있다. 첫번째 분류는 사용자가 부여한 평점을 이용해 추론하는 방법으로서 **AutAI[3]**와 **AdaM[12]**이 그 예이다. 기존 협업 필터링으로 행렬의 빈 셀을 채운다. 그런 다음 채워진 행렬을 이용해 최종적인 추천을 한다. 두번째 분류는 사용자가 부여한 **평점의 존재 유/무를 이용해 평점을 추론하는 방법**으로 **zero-injection[4]**과 **AllRank[6]**가 그 예이다. 사용자가 존재를 알고 있는 두 상품이 있다고 가정해 보자. **만일, 하나의 상품에 대해서는 평점을 부여했고, 하나의 상품에는 평점이 존재하지 않는다면 평점의 높낮이와는 무관하게 해당 상품을 더 선호한다는 것을** 알 수 있다. 평점이 존재하는 상품을 더 선

호하였기 때문에 해당 상품을 이용한 것이고, 이용했기 때문에 만족도를 평점으로 표현할 수 있다고 볼 수 있기 때문이다. **두번째 분류는 비어있는 셀에는 해당 사용자의 부정적인 선호도가 높아있다고 전제한다**. 세부적으로 **전체 비어있는 셀에 사용자의 부정적인 선호도가 높아있다고 바라보는 방법[6,10]**과 **대부분 비어있는 셀에 사용자의 부정적인 선호도가 높아있다고 바라보는 방법[4]**으로 구분된다. **전자는 전체 비어있는 셀에 부정적인 선호도를 나타낼 수 있는 값(0 혹은 2)으로 결정한 뒤 R 에 채운다**. 후자는 **부정적인 선호도를 높아있는 대부분의 비어있는 셀을 찾은 뒤 부정적인 선호도를 나타낼 수 있는 0을 채운다**.

비어있는 셀은 소수의 흥미로운 상품과 다수의 무관심 상품으로 분류할 수 있다. 흥미로운 상품은 상품을 이용하기 전 해당 상품에 흥미를 가졌고, 궁극적으로 사용자가 그 상품을 이용할 것으로 기대한다. 반면, 무관심 상품은 사용자가 해당 상품에 관심이 없어서 이용하지 않을 것으로 기대하는 것이다. 사용자는 시간과 비용의 문제로, 흥미있는 소수의 상품만 이용하게 되고, 대다수의 상품은 이용조차 하지 않는다.

이 관점에서 **첫번째 분류에 속하는 기존 방법들은 흥미로운 상품은 잘 추론할 수 있지만, 무관심 아이템을 그렇지 못하다**. 우선 평점이 존재한다는 것은 사용자가 해당 아이템을 선호했다는 것이고, 이를 이용하면 같은 성격인 흥미로운 상품은 잘 추론할 수 있다, 반면, **무관심 상품은 부정적인 선호도가 높아 있음에도 불구하고, 전체적으로 높은 평점으로 추론하게 되어 오차가 발생하게 된다**. **평점**

두번째 분류는 첫번째 분류와 달리 비어있는 셀에 대해서 낮은 값(0 혹은 2)으로 결정했기 때문에 다수의 무관심 상품의 평점을 잘 추론할 수 있다. 특히, 비어있는 셀 중 다수가 무관심 상품이기에 때문에 첫번째 분류보다 **더 높은 정확도를 보인다**. 그러나 2가지 **한계점이 존재한다**. 첫번째 한계점은 **소수의 흥미로운 상품을 무시하는 것이다**. 전체 비어있는 셀에 대해서 사용자의 부정적인 선호도가 높아 있다고 보는 방법에선, **흥미로운 상품과 무관심 상품의 구분없이 낮은 값을 채운다**. 존재하는 평점의 평균이 높은 것으로 볼 때, 흥미로운 상품의 평점을 낮게 추론하면 오차가 존재하게 된다. **두번째 한계점은 사용자가 부여할 수 없는 평점으로 추론한 것이다**. 무관심 상품은 사용자가 이용조차 하지 않기 때문에 가장 낮은 평점 1 보다 더 낮은 0으로 추론한다. 하지만 추론을 한다는 것은 사용자가 부여할 것 같은 평점을 의미하며, 사용자가 부여할 수 있는 평점 중에서도 부정적인 선호도를 표현할 수 있는 1, 2가 있다. 그렇기 때문에 1보다 더 낮은 0으로 추론하는 것은 사용자의

부정적인 선호도를 더 부정적으로 보게 되어 사용자의 선호도를 왜곡하게 된다. 본 연구에서는 기존 데이터 임putation 방법의 3가지 한계점 실험을 통해 보이고, 어떤 방향으로 데이터 임putation 해야 되는지 대한 아래와 같은 3가지 방법을 제시한다.

- 방법 1: 무관심 상품의 평점은 존재하는 평점을 이용해 추론하는 것이 아니라 부정적인 선호도를 파악할 수 있는 평점 유/무를 이용해 추론해야 한다.
- 방법 2: 흥미로운 상품은 무관심 상품보다 높은 값으로 추론되어야 한다.
- 방법 3: 추론되는 평점은 사용자가 부여할 수 있는 범위 결정해야 한다.

3. 데이터 임putation의 실험 평가 분석

3.1 데이터

추천 시스템 성능 평가에 널리 사용되는 movieLens 100k 데이터를 이용한다[1,5,8]. 이 데이터는 943명의 사용자, 1682개의 상품, 100,000개의 평점으로 구성되어 있다. 평점의 범위는 1(최저)에서 5(최고)이다. 정확도 측정에는 precision을 이용했다. precision의 recall, normalized discounted cumulative gain (nDCG), mean reciprocal rank(MRR)도 이용했으며 모두 같은 경향을 보였다.

3.2 방법 1: 무관심 상품은 부정적인 선호도로 평점 추론

같은 성격인 흥미로운 상품의 평점은 잘 추론할 수 있지만, 무관심 상품의 평점은 부정적인 선호도가 낮아 있음에도 불구하고 높은 값으로 추론되어 오차가 존재한다. 실험은 존재하는 평점을 이용하면 모든 비어있는 셀에 대해서 높은 평점으로 추론되는지 검증한다.

그림 1은 추론된 평점과 존재하는 평점의 누적 분포를 보여준다. 두 분포 모두 3점 이상의 값을 많이 가지

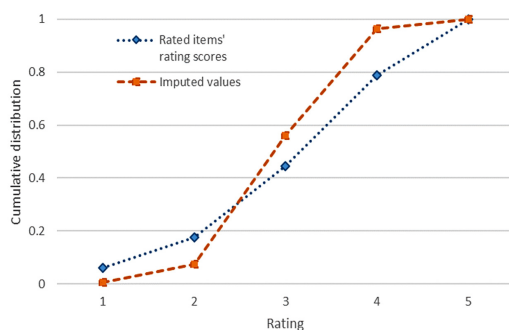


그림 1 추론된 평점과 존재하는 평점의 누적 분포
Fig. 1 Cumulative distribution of I-values of unrated items and existing ratings

고 있으며, 2점 이하의 적다. 추론된 평점의 3점이 이상의 비율이 93%, 2점 이하의 비율이 7%이다. 존재하는 평점을 이용하면 비어있는 셀을 모두 흥미로운 상품으로 전체하기 때문에 전체적으로 유사한 형태의 분포가 된다. 비어있는 셀의 93%가 사용자로부터 관심(3점 이상)을 받고, 그중 44%를 사용자가 좋아했다는 것은 합리적인 상황은 아니다. 특히, 존재하는 평점을 이용하지 평점 유/무를 이용하는 데이터 임putation 방법에서 비어있는 셀의 90%에서 100%를 0 혹은 2로 결정하여, 더 정확한 추천을 한다는 것을 고려하면, 추론된 평점은 전체적으로 너무 높은 편이다. (추론된 평점의 평균 3.39). 이는 비어있는 셀의 대부분이 무관심 상품이라는 것을 보여준다. 위 실험을 통해 무관심 상품을 존재하는 평점으로 추론하면 오차가 크게 발생한다는 것을 확인했고, 이를 해결하기 위해 평점 유/무를 이용해 부정적인 선호도를 반영해 평점을 추론해야 한다는 것을 알 수 있다.

3.3 방법 2: 흥미로운 상품의 평점 추론

기존 방법에서는 사용자가 전체 비어있는 셀에 대해 부정적인 선호도를 가진다는 것을 전제로, 전체 비어있는 셀에 부정적인 선호도를 나타내는 0 또는 2로 결정한다. 평점을 이용한 방법과 비교하면 높은 정확도를 보이는데, 이는 비어있는 셀의 대다수가 무관심 아이템이라는 것을 말한다. 하지만 비어있는 셀 중 소수인 흥미있는 상품의 평점을 고려하지 않고 무관심 상품과 동일한 값으로 결정했다는 한계점을 가지고 있다. 기존 방법 중 흥미있는 상품의 평점과 무관심 상품의 평점을 같은 평점으로 추론하는 방법의 정확도 개선을 통해 방법 2를 검증한다. 기존 방법을 그대로 사용하고, 흥미있는 상품의 평점만 무관심 상품의 평점보다 높은 값으로 결정한다. 흥미로운 아이템은 [4]의 정의를 따라 이용 전 선호도 기준으로 상위 10%를 선택한다.

그림 2는 흥미있는 상품의 추론 평점에 따른 정확도

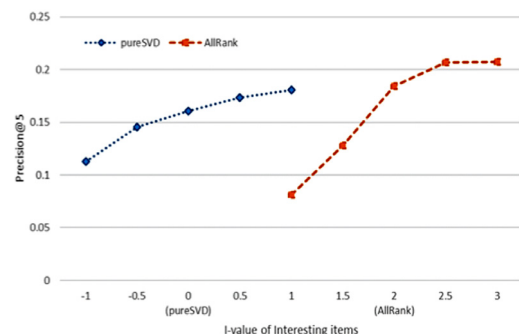


그림 2 흥미있는 상품의 추론 평점에 따른
Fig. 2 The accuracy with different I-values of interesting items

변화를 보여준다. pureSVD[10]는 흥미있는 상품의 평점을 무관심 상품의 평점과 동일한 0으로 결정했다. (AllRanks는 2) 그림 2를 보면, pureSVD, AllRank 모두 흥미있는 상품의 평점을 무관심 상품의 평점보다 높게 결정했을 때 정확도가 상승한다. 또한, 흥미있는 상품의 평점을 무관심 상품의 평점보다 낮게 결정하면 정확도가 떨어진다. 이 실험을 통해 **흥미있는 상품의 평점은 무관심 상품의 평점보다 더 높아야 한다는 것을 알 수 있다.**

3.3 방법 3: 사용자가 부여할 수 있는 평점

기존 방법에서 사용자는 무관심 상품을 추천 받아도 만족하지 않고 이용한 상품을 보다 더 낮은 선호도를 보인다고 보기 때문에 가장 낮은 평점인 1보다 더 낮은 0으로 추론했다[4]. 하지만 추론된 평점은 사용자가 부여할 것 같은 평점을 의미하며, 사용자가 부여할 수 있는 평점 중에서도 만족하지 않음을 나타낼 수 있는 1, 2가 있다. 그렇기 때문에 1보다 더 낮은 0으로 추론하는 것은 사용자의 부정적인 선호도를 더 낮게 보게 되어 추론된 평점에 오차가 존재하게 된다. 기존 방법 중 0으로 추론하는 방법의 정확도 개선을 통해 방법 3의 효과를 검증한다. pureSVD는 전체 비어있는 셀을 0으로 결정하고, Zero-injection[4]은 비어있는 셀 중 무관심 아이템을 찾고 그 셀에만 0으로 결정한다. 실험에선 각 방법을 그대로 사용하고, 0 대신 사용자가 부여할 수 있는 평점으로 결정해 정확도를 비교한다.

그림 3은 사용자가 부여할 수 있는 평점으로 결정했을 때의 정확도 변화를 보여준다. 그림 3을 보면, 두 방법 모두 1, 2, 3에서 기존 방법보다 높은 정확도를 보인다. 특히 3에서 기존 방법 보다 더 높은 정확도를 보이는 것으로 볼 때, 사용자가 부여할 수 없는 평점인 0을 부여한 것은 사용자의 부정적인 선호도를 더 부정적으로 봤다는 것을 알 수 있다. Zero-injection은 2에서 가장

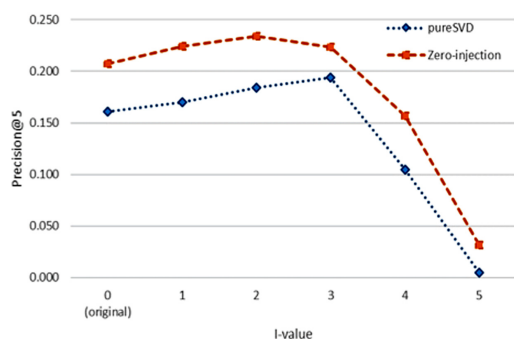


그림 3 추론된 평점의 범위에 따른 정확도

Fig. 3 The accuracy with respect to different I-values ranged between 1and 5)

높은 정확도를 보이는데, 무관심 상품의 평점은 2가 가장 적합하다는 것을 알 수 있다. 한편, pureSVD는 3에서 가장 높은 정확도를 보이는데, zero-injection과 달리 **흥미로운 상품도 채우기 때문이다.** 추가로 4, 5에서 정확도가 급격히 떨어지는 것으로 볼 때, 비어있는 셀의 대부분이 무관심 상품이라는 것을 다시 확인할 수 있다.

4. 결론

최근 상품 수의 증가로 추천 시스템의 중요성이 날로 커지고 있지만, 데이터 희소성 문제는 여전히 추천 시스템의 대표적인 문제로 남아있다. 본 논문에서는 데이터 희소성 문제를 해결하는 기존 방법의 한계점을 정의하고, 이를 개선하는 3가지 방법을 제안했다. 그리고 실험을 통해 한계점을 검증하고, 3가지 방법의 유효성을 보였다.

References

- [1] Jongwuk Lee, Dongwon Lee, Yeon-Chang Lee, Won-Seok Hwang, and Sang-Wook Kim. 2016. Improving the accuracy of top-N recommendation using a preference model. *Information Sciences* 348 (2016), pp. 290-304.
- [2] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. *Proc. of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM*, pp. 426-434.
- [3] Hao Ma, Haixuan Yang, Michael R Lyu, and Irwin King. 2008. Sorec: social recommendation using probabilistic matrix factorization. *Proc. of the 17th ACM conference on Information and knowledge management. ACM*, pp. 931-940.
- [4] Won-Seok Hwang, Juan Parc, Sang-Wook Kim, Jongwuk Lee, and Dongwon Lee. 2016. "Told You I Didn't Like It": Exploiting Uninteresting Items for Effective Collaborative Filtering. *2016 IEEE 32nd International Conference on Data Engineering. IEEE*, pp. 349-360.
- [5] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. *Proc. of the 10th international conference on World Wide Web. ACM*, pp. 285-295.
- [6] Harald Steck. 2010. Training and testing of recommender systems on data missing not at random. *Proc. of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM*, pp. 713-722.
- [7] Mohsen Jamali and Martin Ester. 2009. TrustWalker: a random walk model for combining trust-based and item-based recommendation. *Proc. of the 15th ACM SIGKDD international conference on Knowledge*

- discovery and data mining. ACM*, pp.397-406.
- [8] Won-Seok Hwang, Shaoyu Li, Sang-Wook Kim, and Ho Jin Choi. 2013. Exploiting trustors as well as trustees in trust-based recommendation. *Proc. of the 22nd ACM international conference on Conference on information & knowledge management. ACM*, pp. 1893-1896.
- [9] Jiwoon Ha, Soon-Hyoung Kwon, Sang-Wook Kim, Christos Faloutsos, and Sunju Park. 2012. Top-N recommendation through belief propagation. *Proc. of the 21st ACM international conference on Information and knowledge management. ACM*, pp. 2343-2346.
- [10] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. 2010. Performance of recommender algorithms on top-n recommendation tasks. *Proc. of the fourth ACM conference on Recommender systems. ACM*, pp. 39-46.
- [11] Shuo Chang, F Maxwell Harper, Lingfei He, and Loren G Terveen. 2016. CrowdLens: Experimenting with Crowd-Powered Recommendation and Explanation. 10th International AAAI Conference on Web and Social Media.
- [12] Hao Ma, Irwin King, and Michael R Lyu. 2007. Effective missing data prediction for collaborative filtering. *Proc. of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM*, pp. 39-46.



이 영 남

2017년 한양대학교 컴퓨터·소프트웨어학과 졸업(석사). 2017년~현재 류노 연구원. 관심분야는 추천 시스템, 딥러닝, 의료 정보 분석



김 상 옥

1985년~1989년 서울대학교 컴퓨터공학과 졸업(학사). 1989년~1991년 한국과학기술 전산학과 졸업(석사). 1991년~1994년 한국과학기술원 전산학과 졸업(박사). 1994년~1995년 KAIST 정보전자 연구소 위촉 연구원. 1999년~2000년 IBM T.J.

Watson Research Center, Post-Doc. (invited by Dr. Philip Yu). 1995년~2003년 강원대학교 정보통신공학과 교수. 2003년~2006년 한양대학교 정보통신대학 정보통신학부 부교수. 2006년~현재 한양대학교 정보통신대학 정보통신학부 교수. 관심분야는 데이터베이스시스템, 저장 시스템, 트랜잭션 관리, 데이터 마이닝, 멀티미디어 정보 검색, 공간 데이터베이스/GIS, 주기억장치 데이터베이스, 이동 객체 데이터베이스/텔레매틱스, 사회연결망 분석, 웹 데이터 분석